filosofia unisinos | Unisinos Journal of Philosophy

ISSN 1984-8234

Filosofia Unisinos
Unisinos Journal of Philosophy
25(1): 1-15, 2024 | e25108

Unisinos – doi: 10.4013/fsu.2024.251.08

Article

Intelligence and Philosophy: between new and old artificial crossroads

Inteligência e Filosofia: entre novas e velhas encruzilhadas artificias

Delamar José Volpato Dutra
https://orcid.org/0000-0002-3738-7865
Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Filosofia, Florianópolis, SC, Brasil. Email: djvdutra@yahoo.com.br

Edna Gusmão de Góes Brennand
https://orcid.org/0000-0001-7471-3343
Universidade Federal da Paraíba, Programa de Pós-Graduação em Educação, Paraíba, PB, Brasil. Email: ednabrennand@gmail.com

ABSTRACT

The article discusses which philosophical tools are indispensable and fundamental for understanding the meaning of technology in modern times. It focuses on the debate about artificial intelligence "replacing thought". It provides philosophical criteria for evaluating the rupture between humanism and technology and possible interpretative and analytical choices about human cognition and the possibility of its duplication by machines. It raises ethical and legal issues in the current debates involving Artificial Intelligence and Philosophy, in the sense of the need to regulate the use of artificial intelligence ethically and legally, considering the human perspective of its use and the impacts on humanity. It ponders whether a philosophy of reconstruction of the world will emerge based on technological standards or whether it will look to the humanist tradition to develop an interpretation, given that relations between humans and machines are still unclear.

Keywords: philosophy, artificial intelligence, machine learning.

RESUMO

O artigo discute quais ferramentas filosóficas são indispensáveis e fundamentais para a compreensão do significado da tecnologia nos tempos modernos. Centra-se no debate sobre a inteligência artificial "substituindo o pensamento". Fornece critérios filosóficos para avaliar a ruptura entre humanismo e tecnologia e possíveis escolhas interpretativas e analíticas sobre a cognição humana e a possibilidade de sua duplicação por máquinas. Levanta questões éticas e jurídicas nos debates atuais envolvendo Inteligência Artificial e Filosofia, no sentido da necessidade de regulamentar o uso da inteligência artificial de forma ética e legal, considerando a perspectiva humana de seu uso e os impactos para a humanidade. Pondera se surgirá uma filosofia de reconstrução do mundo baseada em padrões tecnológicos ou se recorrerá à tradição humanista para desenvolver uma interpretação, dado que as relações entre humanos e máquinas ainda não são claras.

Palavras-chave: filosofia, inteligência artificial, aprendizado de máquina.

1 Introduction: machinery and intelligence

The technological evolution and human experience over the last four decades has brought about an important debate between computer science, mathematics and philosophy, regarding human-machine interaction. This topic reached interdisciplinary importance at the beginning of the 1980s and reached its peak at the end of the 1990s. At this point in history, personal computers gain market share, reaching recurrent civilian use and leaving the specialist use zone, expanding the more intensive use of the connection between universities, governments and military bodies. The last two decades of this century have been iconic for the world of technological evolution, and the radical changes brought to educational institutions and companies in the throes of transformation and innovation. The process of globalization of the economy and culture is marked by advances in information systems that accompany trends towards connectivity and global standardization through miniaturization of electronic systems with increased speed and capacity, portability and compatibility of devices and the limitless growth of digitized information. The training of professionals enters society in the so-called data age and reaches an irreversible level of connection between artificial intelligence and social learning. In this process, a new area of knowledge called big data has emerged, whose field of research focuses on how to process, analyze and obtain information from large, complex data sets, from recurring new sources. According to Magrani (2018), the history of the internet can be understood in three generations: the internet of machines, the internet of people and the internet of things.

The process of industrialization introduced, for the first time, intelligent systems capable of carrying out tasks performed by workers. Industrial robots have had an impact on the world of production and work, increasing profitability by improving product quality, reducing production costs, and replacing workers in some tasks. Automation software offers the possibility of performing simple administrative support tasks and is evolving into large-scale automation of more complex processes. In this context, debates are intensifying about the relationship between humanity and technology, and narratives about whether the impacts have been positive or negative. In the fields of exact sciences and nature, humanities and the arts, many questions have arisen about the essence of human beings, fueling the tension between technological evolution and human experience. Since the second decade of this century, the Internet of Things (IoT) and the 5G Internet have brought greater connectivity to all processes involving the use of technology and culture. They have ample data transfer capacity and extremely high connection speed standards with multiple users connected simultaneously. Developed from the idea of a

worldwide network of connected objects with the possibility of exchanging information with each other, the IoT still raises many questions relating to interfaces, design and the user. Studies on the subject, for example by Magrani (2018; 2019), Faccione Filho (2016), Ashton (2010), Atzori et al (2010), show that the number of IoT events and publications is growing. It is possible to verify public interest in the topic by consulting its evolution through Google Insights with the keyword internet of things.

However, there is still a need for studies involving issues of ethics, privacy, legislation, and interoperability, which are important indicators of the breadth of the phenomenon. Various international organizations are carrying out studies and research into the creation of general architectural standards and functionalities for the development of application solutions for the new, large-scale markets that characterize what we can understand as the "thing". The frontiers of discovery are still open. According to Faccioni Filho (2016), this new "network of objects" offers a multitude of new solutions in preparation for a near future that will integrate different technologies and social fields in a diversity of elements and areas of coverage. There is still a need for greater clarity as to the consequences of its social impacts. The prospect is that of connecting an extensive network, with smartphones and industrial robots and other connections that require high performance and structured connectivity. It has established itself as a fluid and necessary ecosystem for tackling major problems around the world. Its use has made headlines in the media all over the planet. One example is the use by academia and science of ChatGPT, an artificial intelligence created by a US research laboratory called OpenAI, whose architecture is based on a neural network called Transformer specially designed to deal with text. The name ChatGPT is the acronym for an algorithm developed using neural networks and machine learning, which focuses on improving virtual dialogues by offering users simple ways to chat and get answers by cross-referencing the data available on the internet. According to Fleck et al (2016), "a neural network is a system designed to model how the brain performs a particular task, implemented using electronic components or by propagation simulation in a digital computer". A neural network is capable of developing learning processes using parameters set by the stimulation of the environment in which it is inserted. In other words, the learning process takes place because, according to Haykin (2001a; 2001b)), its functioning resembles the human brain in two basic aspects: a) knowledge is acquired by the network from its environment, through the learning process; b) connection forces between neurons (synaptic weights) are used to store the acquired knowledge. Thus, it undergoes modifications when it responds to stimuli from the environment, as is the case with the human learning process. If the parameters are free, learning will be creative, and AI is capable of contextualizing facts, writing texts, lyrics, poetry, short stories, programming codes, recipes, etc.

This process of authoring texts has mobilized the international scientific community, which is concerned about the ethical issues surrounding the use of AI. In this context, this article aims to problematize the concept of artificial intelligence and the heated debate about whether it is possible to duplicate human intelligence at the current stage of technological development. According to Childs (2016), we are witnessing the return of the tension between technological advancement and human experience, which is not new. It is a problem that has been discussed with greater emphasis since the 19th century, when the first major reactions against technology were recorded. In the 20th century, authors such as Heidegger (1997), Jacques Ellul (2012a; 2012b.), José Ortega e Gasset (1977) and Carl Mitcham (1989), pointed out criteria for assessing the rupture between humanism and technology, offering indispensable and fundamental philosophical tools for understanding the meaning of technology in modernity, since the transformations have been profound and at an accelerated pace. The debate is raging about the implications for various fields of knowledge of artificial intelligence "replacing thought" and impacting on scientific and technological development, as well as the educational bases for building the future. We understand that the problems of the impact of the technological world on social construction require a multidimensional and interdisciplinary understanding involving analysis by generalists from the social sciences, universalists from philosophy, computer science and the emerging cognitive

sciences. To adequately problematize the issue of technicism versus humanism requires approaches based on multiple fields of knowledge. Thus, in this article, the discussion is based on the double corpus of discourses: the conceptual and the ethical. The methodology used for the study was based on the following descriptors: Artificial Intelligence and Philosophy; Artificial Intelligence; Humanism and Technology; Touring Test; Chinese Room; Internet of Things.

Harari (2020) brings us elements of the Cognitive Revolution of Homo Sapiens. He provides us with a brief history of humanity with important elements for understanding the current Scientific Revolution, such as the most significant inventions of the last century: combustion engine, electricity, nuclear energy, the three-dimensional structure of the DNA molecule, nanotechnology (control of matter on a molecular and atomic scale), pharmaceuticals (molecular modification and principles of optimization of prototype compounds), general advances in research involving molecular biology and genetic engineering, among others. According to Harari (2020), human beings have always been architects of innovation since the cognitive revolution. Humans have created imagined social orders with successive inventions including the agricultural revolution, writing, the invention of currency, the creation of empires, the systematization of science, capitalism, the engine of great inventions. Since the 19th century, new epistemological paths have been under construction in a historiography based on research that has generated anchor points and clarifications on the various historical contexts that have generated the relationship between technology and humanism. Finally, we have arrived at the current 5.0 society, which advocates repositioning technologies for the benefit of human beings. In other words, human beings are at the center of technological transformations. Usage of AI to enhance a social ecosystem where human skills are geared towards promoting inclusion, ensuring sustainability, and improving the quality of human life. Is this a recovery of the Aristotelian vision of the good life (ethics) and the common good (politics) or a recovery of the currents of contemporary practical philosophy (utilitarianism and liberalism)?

Since Touring (1950), the concept of AI has been popularized in many fields of knowledge, with philosophy having an impact of diverse reflections. For methodological purposes, we will present some concepts that are essential for the purposes of this work. According to Koselleck (1992), we understand what a concept is when it is possible to conceive a history through it. This means that not every word can be transformed into a concept. The author argues that, in a simplified way, we can admit that each word refers to a meaning and, consequently, to a content. However, concepts have limitations and boundaries. We will take as our axis the understanding that the concept of AI goes through historical metamorphoses.

That being said, we will look for the concept of Artificial Intelligence on the border between the exact sciences and philosophy, to explain the methodological route that underpins our approach to clarify controversies at the heart of a possible theoretical debate. As a linguistic phenomenon (Habermas, 1968), the concept can transcend the theoretical dimension to act in the understanding of facts in concrete reality. There is no concept without a relationship to a given context. Thus, we searched for the concept of AI in the double corpus: computer science and philosophy. In his text Computing machinery and intelligence, Alain Touring (1950) presents his "game of imitation" in which he shows the roots of the concept of AI, raising a question that still fuels debates among philosophers of the mind: "Can machines think?" This basic question still intrigues philosophers today, since there are many understandings about what kind of intelligence can be produced by algorithms run by electronic artifacts. The so-called Analytical Machine designed by the mathematician Charles Baddage and Ada Lovelace brought about the first theological and philosophical inquiries into the fact that thinking is an exclusive act of the human soul.

Touring (1952) presents seven arguments and nine objections to the so-called Touring Test or The Imitation Game, in which he tests the ability of a machine to produce a narrative in the same way as a human being. In a nutshell, the game consisted of placing three participants in isolated rooms: a man, a woman, and an interrogator, to discern by written answers, without visual or verbal contact, who is a man and who is a woman, based on questions that the interrogator can ask. In the process, a communication channel with a keyboard and a screen is used to generate the result, which today would be a type of

instant messaging tool. One of the interrogated was swapped for the machine and the interrogator had to distinguish between the person and the machine, which didn't happen. For Touring, it would be difficult for a person to imitate a machine, above all because of its inability to perform complex calculations. A machine can perform, for example, billions of calculations and carry out tasks with small margins of error. This ability is very difficult to prove with human beings. Given the various possible entrances to the discussion on this subject in the field of philosophy, we consider the studies carried out by Jonh Searle (1980) as an entry rhizome, when he substantiated the criticisms of Touring's concept of AI, through his Chinese room argument. Searle (1980;2007) published the article *Minds, brains, and programs*, which had great academic repercussions. In his critique, he analyzes what he calls formal processes in which the machine performs tasks based on information sent to it. He claims that even if the answers are positive, the machine is still incapable of understanding the information it is dealing with. The basis of Searle's critique is to challenge Touring's arguments that machines can think, believing that machines are only capable of simulation, have no intentionality and that it is impossible to duplicate the human mind. Searle's (1996) famous example of the Chinese room provides evidence that machines can be programmed to perform formal manipulation of symbols, simulate conversations and responses, but they do not have cognitive states like humans. Humans are different. The author raises the assumptions that programs are totally syntactic, and minds have a semantic capacity: "Minds are semantical in the sense that they have more than a formal structure, they have a content" (Searle, 1997, p. 38-9). The fact of projecting simulations does not guarantee that machines have a mind. Searle's syntactic abilities involve the concept of intentionality, i.e., the ability of the mind to represent objects, understand meanings, present beliefs, and desires in relation to meaning, which is not purely formal. Intentionality can be both original (only humans and animals can desire) and derived (written descriptions and sketches). Dennet (2022) criticizes Searle's concept of intentionality by arguing that if an AI is equipped with mechanisms that allow it to interact with the environment, it can learn through interaction. In broader terms, Searle's critique targets the concept of Artificial Intelligence, which he calls Strong Artificial Intelligence - SAI. Roughly speaking, SAI was the concept coined by Searle to describe a field of understanding in which it is possible to simulate the human mind using a computer model. For the author, consciousness and thinking are biological and uniquely human phenomena that cannot be duplicated, but only simulated. What Searle sought was an attempt to abstractly have an experience, based on his own theory, to elucidate the question posed by Turing: "Can machines think?"

Many objections to Searle's Chinese room test have been raised by various authors of analytic philosophy such as Roger Penrose (2007), Stevan Harnard (2007), Mark Bishop (2007), Copeland (1993, 2007), among others. These criticisms were both of a logical-formal nature and of the author's understanding of fundamental concepts and precepts in the fields of computing and cognitive sciences, pointing to Searle's misunderstanding of Touring's claims, above all because he confused simulation with duplication.

Since Touring and Searle, encouraging advances in technology have brought significant elements into philosophical discussion about the nature of the mind, offering various possible ways of enriching this question. Putnan (1967; 2023) already raised the possible association between Touring's machine and the human being. Authors such as Fodor (1991; 2004), Dreyfus (1992, 1972) and Churchland (1990;2004) offer critics possible alternatives to clarify the problem of what types of evidence are still needed to justify an analytical interpretative choice about human cognition and the possibility of its duplication by machines.

It is also important to note that these questions are being analyzed in the field of culture. Literary and fictional works such as Mary Shelley's novel *Frankenstein* and Aldous Huxley's *Brave New World* are well-known narratives that express the fear of scientific progress in which human relationships are overtaken by technological development. The authors created dystopian representations of the world with fears about the visible consequences of the relationship between technology and humanism. In the cinema, the plot

and philosophical argument of *Blade Runner* (Ridley Scott, 1982), *Matrix* saga (Lana and Lilly Wachowsky, 1999-2003), *Terminator* (James Cameron, 1994) and other films in the genre such as *I Robot* (Alex Proyas, 2004), *Ex-Machina* (Alex Garland, 2015), *Alita - Combat Angel* (Robert Rodriguez, 2019), *M3gan* (Gerard Jonhstone, 2023), show philosophical reflections that have generated debates and diverse theses in an attempt to answer the multiple questions raised about human-machine relations.

2 Human Intelligence and Artificial Intelligence: coming together and moving apart

Multiple questions raised by philosophers, linguists, and sociologists, for example, take the concept of intelligence as an important parameter. Considered complex, it is still in full development. It varies according to different fields of knowledge, for example: psychology (ability to learn and relate), biology (ability to adapt to new habitats or situations). The debate is open and expanding. Materialist conceptions, for example, deny the possibility of reducing the mind to matter. Thomas Nagel (1974) in his article "What is it like to be a bat", in the contemporary philosophy of mind debate, raises the thesis that every mental event or phenomenon can be reduced to a physical event or phenomenon. Nagel denies the irreducibility of the subjective to the objective, in other words, he denies the possibility of explaining our consciousness based on physical phenomena. He maintains that it is impossible to explain the mind from the body. At stake is the definition of what thought is. The project to create an AI that thinks is a purely behaviorist experiment. Many objections to this question are important so that we can follow the research into understanding that a machine can behave similarly to a human being. It can be admitted that this machine can think or have an interior life. The concept of artificial intelligence is complex. Among many approaches, the European Union document, *Ethical Guidelines for Trustworthy AI*, considers that an AI has the capacity to make decisions and adapt its behavior (European Commission, 2019, p. 47).

According to Nicolelis (2023), AI is not intelligent, since intelligence has to do with solving a specific problem, namely the survival of biological species. Therefore, intelligence is characterized by a type of learning specific to biological organisms to survive. Apparently, this wouldn't be a problem for a machine, or at least it would be something they wouldn't yet have in mind, hence the conclusion that it couldn't be intelligent. Nicolelis (2023) presupposes a concept of intelligence and learning derived from a biological view of the phenomenon, but the subject can be dealt with from a more general point of view, such as that according to which intelligence is a human capacity to process and produce information to solve problems and learning is a modification resulting from a stimulus from the environment, in a more general way. According to this view, learning can mean the ability to dynamically adapt behavior, in a non-deterministic way and susceptible to unexpected behavior (European Commission, 2019, p. 26). On the other hand, Gardner (2000, p. 250) defines intelligence as the human capacity to process and produce information to solve problems or produce products that are important to a given culture and cannot be measured. It is the human potential to process information and can be activated or not depending on the cultural configuration. It is made up of the set of skills an individual has for solving problems. If we accept that machines can solve complex problems not solved by human beings, we understand that this debate has only just begun. Many experiments still need to be carried out to make other interpretations possible. In the next section we will gain a better understanding of what is now known as machine learning.

3 Neural networks and machine learning

It is through Machine Learning that computers are acquiring new skills. Machine Learning techniques allow the computer to learn by example, in other words, to learn from data. Machine Learning

has become key to putting knowledge into computers. Humans have a lot of intuitive knowledge, which they can't easily express verbally. We don't have conscious access to this intuitive knowledge. Without a formal understanding of this intuitive knowledge, it is not possible to write programs to represent it. So, what's the solution? The solution is for the machine to learn this knowledge for itself, in a similar way to how human beings learn. In the last two years we have seen a return to the debate on the question posed by Touring (1950; 1952): can machines think? To broaden this debate, we ask: can machines learn?

Currently it is impossible to talk about AI without referring to a system designed to model how the human brain performs a particular task. In its most general form, a neural network is implemented using electronic components or is simulated by propagation in a digital computer. To achieve good performance, neural networks employ a massive interconnection of simple computational cells, called "neurons" or processing units (Haykin, 2001a; 2001b). One of the main and intriguing features is the ability to learn from examples and to generalize the information learned. According to Haykin (2001a), the neural network resembles the human brain in two basic aspects: a) knowledge is acquired by the network from its environment through the learning process; b) connection forces between neurons (synaptic weights) are used to store the acquired knowledge. The learning of a neural network is a process in which the free parameters are adapted through a process of stimulation by the environment in which the network is inserted. The type of learning is determined based on the way in which the parameters are modified. In summary, there is the following sequence of events: a) the neural network is stimulated by an environment; b) the neural network undergoes modifications to its free parameters because of this stimulation; c) the neural network responds in a new way to the environment, due to the modifications made to its internal structure (Haykin, 2001b).

Artificial Neural Networks (ANN) are mathematical models that are inspired by biological neural structures and whose computational capacity is acquired through learning. The processing of information in ANNs is done in artificial neurons, known as McCulloch e Pitts, 1943; 1990). Many mathematical models are inspired by biological neural structures. These models show that it is possible to design and increase the computational capacity acquired through learning. To learn a neural network, the algorithm needs to go through the training and testing phases where the parameter settings are modified and evaluated. Neural Networks no longer have fully connected structures and are inspired by the local sensitivity and selective orientation of the brain. They are currently solving most problems satisfactorily. They somehow summarize the activations of the neurons they connect to. The great evolution of AI applications in various fields of knowledge and the repercussions of this on reflections on its human implications show that this is a developing phenomenon and promises to amplify studies of possible dysfunctions in applications in society. The forms of use and the necessary regulations are in progress. Below are some ethical and legal issues in the debates involving AI.

4 Current issues: legal normativity applied to AI

So far, we have tried to address philosophical debates and conceptual clarifications about whether artificial intelligence and human intelligence can be considered correlated, as well as whether machines can think or only simulate human learning. We are convinced that there is still a lot of research to be done into the consequences of machine learning in the context of current scientific and technological advances. The debate is not new, but in the last decade it has become more urgent. Therefore, we will now look at the other essential aspect to be considered in this discussion, which is normative issues. At least two paths open in this regard.

[I] The first refers to how robots or AIs should be treated by humans, despite the certainty of being able to predicate them something like intelligence, learning, consciousness, will, intentionality, freedom. This could be done based on the concept of person (Dall'Agnol, 2020), which dispenses these qualities,

as already practiced by our normative systems in relation to babies or human beings with dementia. This perspective can also be extended to artificial agents, as has already been done for corporate legal entities in the legal field. In this sense, it is worth analyzing whether objectives such as those envisaged by Hauert (2023, p 47) are ethical: "Robots are not going to replace humans, they are going to make their jobs much more human. Difficult, humiliating, demanding, dangerous and monotonous - these are the jobs that robots will do".

[II] The second path is one that seeks to regulate the use of AI ethically and legally, with a human perspective in mind. This perspective is based on the massive use that is already being made of AI, and which tends to grow, a use with impacts for humanity. For this perspective there is no need to have a deep understanding of the phenomenon of artificial intelligence, either in the sense of knowing whether it is intelligent, or whether it can learn, since a more pragmatic perspective can be taken, given that there is already sufficient evidence of important impacts to demand normativity. The European Union document, *Ethical Guidelines for Trustworthy AI* (2019), lists concerns of high impact and uncertainty, also lists concerns regarding risks (European Commission, 2019, p. 43s): the possibility of involuntary, non-consensual and mass identification, as well as the possibility of localization; the development of human-like robots, which can generate emotional bonds; the possibility of classifying people in all aspects, creating registers; the creation of autonomous lethal weapons.

As has been said, this path takes shapes from the human perspective, something that can be seen in the research carried out by MIT's Moral Machine website [https://www.moralmachine.net/hl/pt], whose aim is "gathering a human perspective on moral decisions made by machine intelligence". This led Savulescu, Gyngell & Kahane (2021) to propose a procedure for how to use such preferences in political deliberations, including the consideration of what he calls robust philosophical agreements on certain ethical issues. Let's look at examples of concerns in this direction: a) the use of human brain tissue in AI, i.e., kinds of semi-biological devices: "In practice, this means that semi-biological semiconductors will be able to have what Razi calls 'continuous lifelong learning' - which allows the chips to acquire new capabilities without abandoning old learned skills" (Estadão, 2023); b) the incorporation of language models that allow for something like an artificial brain, which boosts learning possibilities (Roose, 2023).

There is also concern about the use of AI in various sectors, with significant impacts: the use of robots in care services (Schaeffer, 2023); the use of robots in the public service, such as granting social security benefits (Gercina, 2023); the use of AI in tax jurisdiction (Salles, 2023); the prediction that artificial neural networks will have a significant impact on cultural production (SENAI, 2023, p. 10-13; Palmeira, 2023); the unemployment that AI can cause (BBC, 2023); the use of AI in credit and insurance sectors, which can lead to predictions being a kind of self-fulfilling prophecy (Jobim do Amaral, Santos Elesbão, da Veiga Dias, 2022, p. 689), thus giving this tool an unprecedented role. The concept of 'machine' helps to understand the phenomenon, since machines can perform complex activities without any human intervention, such as driving cars, flying airplanes, diagnosing diseases, responding to threats. Leben (2019, p. 1) understands that, in this sense, such machines are no more automatic, but autonomous.

These new developments in technology and new possibilities for the application of AI have generated studies in various areas, as well in the field of normativity, whether legal (Branco, Tefé, 2023) or ethical: "the use of robots by the judiciary requires active human vigilance", because "there is no way to attribute malice or guilt to AI" (CONJUR, 2023); AI needs to be controlled (Hinton, 2023); the use of AI can cause harm (Rodas, 2023); Facebook was ordered to pay twenty million reais in moral damages (Mendes, 2023).

Based on the concepts of unacceptable risk, high risk and minimal risk, as well as potential danger, on June 14, 2013, the European Union approved regulations concerning AI. According to this legal regulation, "The development and use of systems that present an unacceptable risk are prohibited, while high-risk systems are subject to severe restrictions on development, implementation and use. Regarding low-risk systems, the requirements relate to transparency" (Sarlet, 2021, p. 298).

Specifically, concerning the Brazilian case, from a legal point of view, the Constitution can play an important role in regulating the matter, due to its deeply principled nature, which gives it a more extended applicability. In the words of the commentator,

> *"It is recognized that, although the 1988 Federal Constitution is not exactly digital, Brazil already has a legal system anchored in principles and provisions that guarantee the enjoyment of rights such as intimacy, equality, freedom, security, non-discrimination and the free development of the personality which, in themselves, already consolidate a significant support for opposing resistance to some of the deviations mentioned above" (Sarlet, 2021, p. 290-1).*

In addition, it should not be forgotten that in 2018 the *General Law on the Protection of Personal Data* came into force, which, although it does not deal directly with AI, has applicability to aspects involved in the issue. This law, called the Constitution of the Internet in Brazil (2018), lists various human rights that must be respected, and created the National Council for the Protection of Personal Data and Privacy. In any case, the author concludes that "based on constitutional principles, the catalog of fundamental rights and guarantees, as well as the rights extracted from the Brazilian legal system, aligned in a normative constellation by human rights, it becomes possible to act positively in this area" (Sarlet, 2021, p. 294).

5 Ethical normativity applied to AI

Ethics, like law, must operate Based on the possibility of harm or risk of harm, which puts humans in the position of victims, being this perspective defended, for example, by Leben (2019), who mentions Asimov's three laws of robotics, the first of which is precisely not to cause harm to a human being (p. 1-2). It is in this sense that studies on prejudice in the uses of AI can be understood, such as in the analysis of credit and insurance profiles (Jobim do Amaral, Santos Elesbão, da Veiga Dias, 2022, p. 689), as well in relation to gender, race and diversity (Barbosa, Tresca, Lasuchner, 2023). With this theme in mind, in 2023 the journal *Res publica* launched a special issue (V. 29) on algorithmic discrimination and fairness, the purpose of which was to address the normativity arising from the moral notion of equality and its relationship with the use of AI.

The European Union document *Ethical Guidelines for Trustworthy* AI (2019) is a good example of how a precautionary principle can be used to deal with AI from an ethical point of view. The document aims to "Develop, deploy and use AI systems in a way that adheres to the ethical principles of: respect for human autonomy, prevention of harm, fairness and explicability" (European Commission, 2019, p. 2). This document has characteristics that allow to take it as a model, since it is positioned at the interface between academic-philosophical studies in the area of applied ethics and deliberative processes for establishing specific ethical standards concerning AI, for the bloc of European Union countries: "these Guidelines seek to go beyond a list of ethical principles, by providing guidance on how such principles can be operationalised in sociotechnical systems" (European Commission, 2019, p. 2). The document is an explanation of the precautionary principle, since it makes extensive use of notions as risk and harm, including the concepts of unacceptable or high risk, which have already been reflected in European Union legislation in 2023, as pointed out above, with the aim of an "ethical, secure and cutting-edge AI" (European Commission, 2019, p. 5).

The precautionary principle is widely used in environmental law (Freitas, 2003). Arguably, its ancestor is the principle of responsibility proposed by Jonas. According to him, an ethics for a technological civilization require a principle of responsibility, the basis of which, he claims, is fear: "We know much sooner what we do not want than what we want. Therefore, moral philosophy must consult our fears prior to our wishes to learn what we really cherish" (Jonas, 2006, p. 71). Ethics based on risk, and therefore on precaution or

prevention, is strongly driven by fear, something that Hobbes dealt with exemplary grandeur in his political philosophy. By the way, in the movie *Terminator*, 1984, the fear of machines is iconic.

Be that as it may, alongside the benefits, the document considers risks or impacts concerning the use of AI "difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice, or on the human mind itself)" (European Commission, 2019, p. 2). The document, as the term itself indicates, in view of the diagnosis made, points out seven key requirements that must be met in order to achieve a trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability. These requirements, once met, would guarantee trust, as they would either prevent or reduce risks. The document clearly states: "Our strategy to help Europe realize these benefits is to use ethics as a key pillar to ensure and develop a trustworthy AI." (European Commission, 2019, p. 6). The other two pillars are legal and robustness, the latter understood in the sense of prevention, i.e., "safeguards to avoid unintended negative impacts" (European Commission, 2019, p. 8). In the end, the main objective is to "protect people and groups at the most basic level" (European Commission, 2019, p. 11). In this vein, of the four principles listed, the second is the prevention of harm, alongside respect for human autonomy, fairness and explicability (European Commission, 2019, p. 14).

The last requirement deserves some emphasis. Responsibility is still attributed to human programmers (Zerilli, 2021, p. 70s), but, considering that the concept is problematic in relation to humans themselves, there are attempts to assign responsibility, in some way, also to machines. For example, Floridi & Sanders (2004) distinguish between responsibility and accountability (p. 351). This distinction allows them to explain why a dog that bites someone is punished, for example, by its isolation, or even explain why the dog undergoes training to try to change its aggressive behavior. (Noorman, 2023). This same distinction could allow to attribute moral accountability to machines, but not moral responsibility.

6 Final considerations

By way of conclusion, it is worth asking whether the modern world of big data, remote simulators, drones, the internet of things, 5G technology, and other inventions using AI justifies and ensures a richer and a more humane future. Will a new philosophy trying to reconstruct the world based solely on technological standards emerge, or will we look to the humanist tradition to develop an interpretation of the world where the relationship between humans and machines remains unclear? Case (2023) in an interview to the newspaper *El Pais* argued that technology humanizes us, at the same time as it gives us the status of cyborgs, since it alters our extensions in terms of our ability to speak, hear and move around. She argues that the synergy between information circuits, digital devices, artificial intelligence, communication networks, social movements and individuals, rewrites and resignifies ways of living and ways of building the social, because of the symbiosis between science, technology, and society. The interaction between humans and non-humans brings to the forefront of discussion the concept of the cyborg and the consequent plasticity in human learning through human-machine interaction (Adam, 2007). By announcing that the person-machine relationship makes us all hybrids, that is cyborgs, she draws attention to the fact that we are not Robocop or the Terminator. We don't wear robotic legs or chips, but we are mental cyborgs every time we use digital devices to live in a world that connects the extensions of our ability to see, hear and move, with AI techniques that are increasingly invading the social world to help us solve complex problems. According to Case (2023), we need to create spaces for thinking and living real experiences.

The desire to make AI learn is not new. Ever since Alan Turing's test, we've known that systems can be intelligent. Even today, AI systems undergo these tests to verify their ability to learn. So far, it has not been

established that humans' ability to learn has not yet been achieved. Research into the learning capacity of machines continues to adjust machines to learn like humans. Although we don't yet know for sure how humans learn, there are already efficient algorithms for "teaching" specific tasks to computers. AI is already learning to drive cars (Google and Tesla cars). But there is still a lack of ethical guidelines and legislation to produce tests in real environments and risk situations (Feng et al., 2021). Studies are needed on the social and ethical impacts of AI, as well as studies on its risks and benefits in the development of human actions.

Meanwhile, the debate is permeated by both unfounded fears and real problems. AI can have both good and bad impacts. AI can prevent human beings from being exposed to dangerous tasks that can be carried out by machines without having to stop dealing with exciting and enjoyable creations. The benefits are already numerous: improvements in the field of health; Natural Language Processing; clean energy production; fraud monitoring; safer means of rapid transportation; more productive governance. It must be recognized that AI also has negative impacts: unemployment and increased social inequalities. As pointed out in previous topics, its use involves numerous ethical and moral issues, including: the use of powerful and automatic weapons; invasion of privacy; lack of transparency in the use of data, among others.

It is known that the concept of AI is disputed, both in terms of the predicates that make it up and the scope and limits of its use. This text has attempted to make normative considerations, both from the point of view of how humans should treat AI, and in relation to how humans should protect themselves from possible harm arising from the use of AI. In both perspectives, legal or ethical standards can be established. In the case of the latter, with a view to protecting humans from possible harm, it was considered that the principle underlying the system proposed by the European Union document, Ethical Guidelines for Trustworthy AI, is that of precaution, already widely used in the area of environmental law, whose ancestor is the principle of responsibility, developed by Jonas, as a possible basis for ethics for a technological society.

The interpretation of the European Union document led to the understanding that it is a good proposal for an ethical system applied to AI, based on the precautionary principle. The document proposed a system of rules structured as a precaution in relation to AI, capable of generating trust. For this reason, the driving force behind the document seems to be fear of AI, which generates distrust. Ethical and legal norms can form a basis capable of dispelling this fear, as they can be powerful enough to establish appropriate precautions, the result of which could be a more trusting relationship with AI or suffering when we are exposed to constant connections, which is why human-machine relations need to be redefined. According to Oliveira (2017), the world of culture is being formalized as a second nature produced by human ingenuity, an ecosystem of cybernatives, of hybrid humans, sanctioned by symbiotic hyperconnectivity, in other words, by the incorporation of inorganic as a potentializer of a new existential status, where the main character of this new paradigmatic status is the cyborg. This figure of fiction and reality is made up of a large network of processes involving technology, the body, subjectivity, social, educational, and political developments. According to the author, a philosophy of the cyborg is being built, to think about human learning and machine learning without falling into the human-machine duality. To wbuild a possible diagnosis of the present, we need to think about forms of epistemic cyborgian practices, ontological cyborgian practices and ethical cyborgian practices. According to Oliveira (2017), if for Spinoza (2007) no one has yet determined what a body can do nowadays it would be essential to ask: what can the mind do?

References

ADAM, A. 2002. Cyborgs in the chinese room: boundaries transgressed and boundaries blurred. In: *Views into the chinese room: new essays on Searleand artificial intelligence*. Edited by John Preston and Mark Bishop. New York: OxfordUniversity

AMARAL, A. J.; ELESBÃO, A. C. S.; DIAS, F. V. 2021. Governamentalidade algorítmica e novas práticas

punitivas. In: *Derechos en Acción*, **20**(20): p. 549.

ASHTON, K. 2009. That 'Internet of Things' thing. In: *RFID Journal*. Disponível em: https://www.rfidjournal.com/that-internet-of-things-thing. Acesso em 13 julho de 2023.

ATZORI, L.; IERA, A.; MORABITO, G. 2010. The Internet of Things: a survey. In: *Computer Networks*, **54**(15): p. 2787-2805.

BARBOSA, B.; TRESCA, L.; LASUCHNER, T. 2023. *TIC, Governança da internet, Gênero, Raça e Diversidade - Tendências e Desafios*. São Paulo: Comitê Gestor da internet do Brasil.

BBC. 2023. A empresa que trocou 90% dos funcionários do SAC por inteligência artificial. *BBC News Brasil*. 13/07/23, acesso em 01/08/23. [A empresa que trocou 90% dos funcionários do SAC por inteligência artificial (msn.com)].

BISHOP, M. 2007. Dancing with pixies: strong artificial intelligence and panpsychism. In: *Views into the chinese room*: new essays on Searle and artificial intelligence. Edited by John Preston and Mark Bishop. New York: Oxford University Press.

BRANCO, S.; TEFFÉ, C. S. 2023. *Inteligência artificial e Big Data.* Diálogos da pós-graduação em Direito Digital. Rio de Janeiro: ITS - Instituto de Tecnologia e Sociedade.

BRASIL. 2018. *Lei Geral de Proteção de Dados Pessoais - Lei nº 13.709/2018.* Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: 11 de mar. 2023.

CASE, A. 2017. *O celular é o novo cigarro: se fico entediada, dou uma olhada nele. Está nos escravizando.* Disponível em: <https://brasil.elpais.com/brasil/2017/12/05/tecnologia/1512483985_320115.html>. Acesso em: 02 de setembro de 2023.

CHILDS, P. 2017. *Modernism***.** New York: Oxon.

CHURCHLAND, P. M. 2004. *Matéria e Consciência: Uma introdução contemporânea à filosofia da mente*. São Paulo: Editora UNESP.

CHURCHLAND, P. M.; CHURCHLAND, P. S. 1990. Could a machine think? *Scientific American*, **262**(1): p. 32-39.

COMISSÃO EUROPEIA. 2019. *Orientações éticas para uma IA de confiança*. GPAN. Comissão Europeia. Bruxelas.

COPELAND, K. 1993. *Artificial intelligence: a philosophical introduction.* Blackwell Publishing.

COPELAND, K. 2007. *The Chinese room from a logical point of view.* New York: Oxford University Press.

DALL'AGNOL, D. 2020. Human and Nonhuman Rights. *Revista de Filosofia Aurora*. **32**(55): p. 4-26.

DENNETT. D. C.; KINSBOURNE, M. 2016. *O tempo e o observador: o onde e o quando da consciência no cérebro.* Disponível em: <http://criticanarede.com/docs/tempo.pdf>. Acesso em: Janeiro de 2023.

DREYFUS, H. 1992. *What Computers Still Can't Do: a critique of artificial reason*. Cambridge: MIT Press.

ELLUL, J. 2012a. *Le système technician. Paris: Le cherche midi*.

ELLUL, J. 2012b. *Le bluff technologique*. Paris: Fayard.

ESTADÃO. *Computador quase humano: conheça o chip com tecido cerebral que promete mudar a IA.* Disponível em: <https://www.estadao.com.br/link/gadgets/chip-tecido-cerebro-humano/>. Acesso em: 09 de outubro de 2023.

FACCIONI F. M. 2016. *Internet das coisas.* Palhoça: UnisulVirtual.

FLECK, L.; TAVARES, M. H. F.; EYNG, E.; HELMANN, A. C.; ANDRADE, M. A. M. 2016. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, **1**(13): p. 47-57.

FLORIDI, L.; SANDERS, J. W. 2004. On the Morality of Artificial Agents. *Minds and Machines*. **14**(3): p. 349-379.

FLORIDI, L.; SANDERS, J. W. 2004. On the morality of artificial agents. In: *Minds and Machines*. **14**(3): p. 349-379.

FODOR, J. 1991. You can Fool Some of the People All of the Time, Everything Else Being Equal. "Hedged Laws and Psychological Explanations". In: *Mind*, **100**(397): p. 19-34.

FODOR, J. 2004. The Mind-Body Problem. In: HEIL, J. *Philosophy of Mind: A Guide and Anthology*. Oxford university Press, p. 168-182.

GARDNER, H. 2000. *Inteligência: um conceito reformulado.* Trad. Adalgisa Campos da Silva. Rio de Janeiro: Objetiva.

GERCINA, C. 2023. INSS aumenta análise de aposentadorias por robôs e nega benefícios em seis minutos. In: *Folha de São Paulo*. Disponivel em: https://www1.folha.uol.com.br/mercado/2023/07/inss-aumenta-analise-de-aposentadorias-por-robos-e-nega-beneficio-em-seis-minutos.shtml. Acesso em: 01 de agosto de 2023.

HABERMAS, J. 1968. *La technique et la science comme ideologie*. Paris: Gallimard, 1968.

HARARI, Y. N. 2020. *Sapiens: uma breve história da Humanidade*. Porto Alegre: Campanhia das Letras.

HARNAD, S. 2007. *Minds, machines, and Searle 2: what's right and wrong about the chinese room argument.* New york: Oxford University Press.

HAYKIN, S. 2001a. Kalman filters. In: *Kalman filtering and neural networks*, John Wiley & Sons. p. 1-21.

HAYKIN, S. 2001b. *Redes Neurais: Princípios e Práticas*. BOOKMAN, São Paulo, 2ª ed. 900 p.

HEIDEGGER, M. 1977. *The Question concerning technology*. Nova York: Harper.

HINTON, G. 2023. *Se existe alguma maneira de controlar a inteligência artificial, devemos descobri-la antes que seja tarde demais*. Entrevista com Geoffrey Hinton - Instituto Humanitas Unisinos - IHU. 09/05/23, acesso em 01/08/23.

JONAS, H. 2006. *O princípio da responsabilidade: ensaio de uma ética para a civilização tecnológica*. Rio de Janeiro: Editora PUC-Rio.

KOSELLECK, R. 1992. Uma história dos conceitos: problemas teóricos e práticos. In: *Revista Estudos Históricos*, **5**(10): p. 134-146.

LEBEN, D. 2018. *Ethics for Robots: How do Design a Moral Algorithm*. Routledge.

MAGRANI, E. 2018. *A internet das coisas*. Rio de Janeiro : FGV Editora.

MAGRANI, E. 2019. *Entre dados e robôs: ética e privacidade na era hiperconectividade*. Porto Alegre, RS: Arquipélago.

MALLMANN, D. MAGALHÃES, T.; BERNARDES, V. 2023. *Facebook é condenado a pagar R$ 20 milhões por vazamento de dados de usuários*. Disponível em: <https://www.cnnbrasil.com.br/nacional/facebook-e-condenado-a-pagar-r-20-milhoes-por-vazamento-de-dados-de-usuarios/>. Acesso em: 01 de agosto de 2023.

MCCULLOCH, W. S.; PITTS, W. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, **5**: p. 115-133, 1943.

MCCULLOCH, W. S.; PITTS, W. H. 1990. A logical calculus of the ideas immanent in nervous activity. In: *Bulletin of mathematical biophysics*. p. 99-115.

MELO, J. O. 2023. *Inteligência Artificial vai revolucionar advocacia, diz consultor dos Estados Unidos*. Disponivel em: <ConJur - IA vai revolucionar a advocacia, diz consultor dos Estados Unidos>. Acesso em: 05 de agosto de 2023.

MENDES, D. 2023. Justiça de MG condena Facebook em R$ 20 milhões por vazamento de dados; veja como pedir indenização. In: *CNN Brasil*. Disponível em : https://www.cnnbrasil.com.br/economia/

justica-de-mg-condena-facebook-em-r-20-milhoes-por-vazamento-de-dados-veja-como-pedir-indenizacao/, acesso em 03/08/23.

MITCHAM, C. 1989. *Qué es la filosofía de la tecnología?*. Barcelona: Editorial Anthropos.

NAGEL, T. 1974. "What is it like to be a bat?". *The Philosophical Review*, **LXXXIII** (4): p. 435-450.

NICOLELIS, M. 2023. *Inteligência Artificial: tudo o que você precisa saber - Miguel Nicolelis*. 12 de junho de 2023. Youtube. Disponível em: https://www.youtube.com/watch?v=pb4b4_MlNwo. Acesso em 01 de agosto de 2023.

NOORMAN, M. 2023. Computing and Moral Responsibility. *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & U. Nodelman (eds.), URL = https://plato.stanford.edu/archives/spr2023/entries/computing-responsibility/, acesso em 02/08/23.

OLIVEIRA, D. F. 2017. *Sobre Humanos e Máquinas: marcos epistêmicos, ontológicos e éticos para compreensão do ciborgue e a aprendizagem humana na cultura digital*. 2017.269 f. (doutorado em Educação) – Universidade Federal da Paraíba, João Pessoa.

ORTEGA Y GASSET, J. 1977. *Meditación de la **técnica: y** otros ensayos*. 7. ed. Madrid: Revista de Occidente.

PALMEIRA, C. 2023. Meta lança AudioCraft, IA generativa que cria músicas e sons a partir de textos. *Tecmundo*. Disponível em: https://www.tecmundo.com.br/software/267126-meta-lanca-audiocraft-ia-generativa-cria-musicas-partir-textos.html. Acesso em: 03 de agosto de 2023.

PENROSE, R. 2007. *Consciousness, computation, and the chinese room*. New York: Oxford University Press.

PUTNAM, H. 1967. The nature of mental states. *Art, Mind, and Religion*. Pittsburgh, University of Pittsburgh Press.

PUTNAM, H. 1975. The Mental Life of Some Machines. In: *Mind, Language and Reality*. Cambridge: Cambridge University Press, pp. 408-428.

RODAS, S. 2023. Lei deve focar na prevenção a danos da IA, não só na responsabilização posterior. In: *Revista Consultor Jurídico*. Disponível em: <https://www.conjur.com.br/2023-mar-21/lei-focar-prevencao-danos-ia-nao-responsabilizacao>. Acesso em: 01 de agosto de 2023.

ROOSE, K. 2023. Revolução na robótica fica mais próxima com IA do Google. In: *Estadão*. Disponível em: https://www.msn.com/pt-br/noticias/ciencia-e-tecnologia/revolu%C3%A7%C3%A3o-na-rob%C3%B3tica-fica-mais-pr%C3%B3xima-com-ia-do-google/ar-AA1eDj12?ocid=msedgntp&cvid=55ff93c895ae4f7a8262182dd617842c&ei=. Acesso em: 01 de agosto de 2023.

SALES, O. 2023. Sistemas de IA estão desenvolvendo habilidades imprevisíveis e cientistas não sabem os motivos. *Estadão*. Disponivel em: https://www.msn.com/pt-br/noticias/ciencia-e-tecnologia/sistemas-de-ia-est%C3%A3o-desenvolvendo-habilidades-imprevis%C3%ADveis-e-cientistas-n%C3%A3o-sabem-os-motivos/ar-AA1d0iwj?ocid=msedgdhp&pc=U531&cvid=a0615f0b561543c69dba3e3841175311&ei=19#image=1. Acesso em: 25 de junho de 2023.

SALLES, L. C. 2023. *Peculiaridades do direito processual do trabalho: inexistência de rito sumário, descabimento da expressão'procedimento sumaríssimo'e unicidade do rito ordinário*. Disponível em: https://repositorio.ufu.br/handle/123456789/37115.

SARLET, G. B. S. 2021. A Inteligência Artificial no contexto atual: uma análise à luz das neurociências voltada para uma proposta de emolduramento ético e jurídico. *Revista Direito Público*. **18**: p. 273-305.

SAVULESCU, J.; GYNGELL, C.; KAHANE, G. 2021. Collective Reflective Equilibrium in Practice (CREP) and Controversial Novel Technologies. *Bioethics*. **35**: p. 652-663.

SCHAEFFER, C. 2023. China terá 100 robôs enfermeiros em teste até o final do ano. *Tecmasters*. 29/07/23, acesso em 01/08/23. [China terá 100 robôs enfermeiros em teste até o fim do ano; conheça o GR-1 | TecMasters].

SEARLE, J. R. 1996. Mentes, cérebros e programas. In: TEIXEIRA, J. F. (Org.), *Cérebros, Máquinas e Consciência: Uma introdução à Filosofia da Mente*, São Carlos: Editora da UFSCar, pp. 61-93.

SEARLE, J. R. 1997. *A Redescoberta da Mente*. São Paulo: Martins Fontes.

SEARLE, J. R. 2007. *Twenty-one years in the chinese room.* New York: Oxford University Press.

SENAI. Departamento Regional do Paraná. 2023. *Tendências Sistema FIEP 2023*. Curitiba: SENAI/PR.

SPINOZA, B. 2007. Theological-Political Treatise. London: Cambridge University Press.

TURING, A. M. 1950. Computing Machinery and Intelligence. *Mind*, **LIX**(236): p. 433-460.

TURING, A. M. 1952. The Chemical Basis of Morphogenesis. In: *The Royal Society.* **237**(641): p. 37-72.

YAN, X.; FENG, S.; SUN, H.; LIU, H. X. 2021. *Distributionally Consistent Simulation of Naturalistic Driving Environment for Autonomous Vehicle Testing.* Disponível em https://arxiv.org/abs/2101.02828. Acesso maio de 2021.

ZERILLI, J. 2021. *A citizen's guide to artificial intelligence*. MIT Press, 2021.

Submetido em 24 de outubro de 2023.

Aceito em 11 de janeiro de 2024.